

Standard, Meta-Standard: a framework for coding occupational data

Greenstein, Daniel I.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Greenstein, D. I. (1991). Standard, Meta-Standard: a framework for coding occupational data. *Historical Social Research*, 16(1), 3-22. <https://doi.org/10.12759/hsr.16.1991.1.3-22>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Standard, Meta-Standard: A Framework for Coding Occupational Data

*Daniel I. Greenstein**

Abstract: Debate over appropriate schema for coding occupational data has been ongoing without satisfactory resolution since at least the late nineteenth century. It is fuelled by the fact that classifying occupational data whether they are collected by culling the historical record or through precise sociological survey, can never be exact. Some of the relevant data are nearly always ambiguous (when is a »merchant« merely a small shop-keeper and when a multi-national shipper of luxury goods?). Moreover, any scheme will inevitably reflect its author's particular research interests and/or assumptions about social and occupational structures relevant to the period and place under investigation. Consequently, any two authors faced with the same dataset are likely to produce different and even incompatible coding schemes with which to categorize occupational information. Authors concerned with similar phenomena which occur in different places or at different times are even more likely to generate incompatible schemes. These well-known problems have far reaching ramifications. If coding schemes are ultimately subjective, then can we ever truly verify the quantitative historical research which employ them? How, without some attempt to standardized coding practices, the quantitative data collected and computerized by others will never be usable in for secondary or comparative analyses?

The following paper addresses these problems head on and suggests that in some respects they are intractable. But the outlook needn't be quite so pessimistic. The usefulness of standard typologies, the paper implies, shouldn't necessarily be measured by the extent to which their categories are adopted elsewhere, but through the general applicability of the con-

* Address all communications to: Daniel I. Greenstein, Corpus Christi College, Oxford, OX1 4JF, Great Britain.

ceptual apparatus used in their construction. If historians can agree how to go about constructing their coding scheme and, as importantly, about how to document concisely the subjective inferences they are employing when composing them, then their work may be genuinely accessible to others. What is necessary, may be a framework within which coding exercises should be conducted; a meta-standard rather than a standard. The paper then makes an initial attempt to develop such a framework with reference to work that has been conducted on a large-scale prosopographical study of Oxford University's twentieth century members. Preparing occupational data for meaningful analysis is infuriatingly difficult. Even data gathered through the most carefully designed survey questionnaire can be ambiguous and even misleading. The problem facing the historian who is dependent on an incomplete written record is even greater. What, for example, does one do with the »merchant« who appears with such irksome regularity in late-nineteenth-century city directories? Was he merely a small shopkeeper or the owner of a large conglomerate enterprise? Too frequent reliance on the »not codeable« category for such unreliable data quickly narrows the analytical horizons of even the largest samples. Introducing probabilities to treat such data only recognizes the problem and provides an elegant defense against would-be critics; it does not solve it.

The problems involved in categorizing occupational information are only compounded by the fact that any scheme is essentially contestable, reflecting at some level the assumptions and particular research interests of its author. They are also not entirely new. Preparation of US Census manuscripts in the late-nineteenth-century generated at least some comment in statistical journals of the day where the merits and utility of various schemes were contested. In academic circles, discussion of the problems involved in the analysis of occupational information awaited the maturation of sociology with the social surveys of David Glass (1) in the United Kingdom and Paul Lazarsfeld in the United States.(2) In history there wasn't much to discuss until well into the 1950s when a critical mass of historical demographers and social historians found themselves engaged in an active search for the common folk with the fragmentary evidence available in a voluminous literature comprising parish registers, census manuscripts and city directories.(3) By the mid-1960s what little discussion was available in learned historical journals suggested that the interpretive problems were already recognized and, though never admitted outwardly perhaps, seen as intractable.(4) Occupational data were quite simply incomplete; their analysis fraught with pitfalls.

There was, however, no call to retreat. The sources of quantitative social history, despite their manifold problems, promised and, I think, provided new insights into the past when treated sensitively and with due regard to their limitations. A researcher was encouraged in the methodological li-

terature that began to emerge from the mid-1960s to approach the census, for example, only when the information it contained, with all its limitations, promised to shed some light on the questions for which answers were sought. Having settled on the census as a relevant source, it was necessary to peer into the world of the enumerator to comprehend the meaning of the descriptive terminology used to indicate people's employment. Coding data once collected was equally driven by this problem-guided and source-sensitive approach. Schemes were devised to answer the particular questions that the researcher set out with.(5)

Tightly even narrowly defined coding schemes had additional advantages as well. They could be accommodated, for example, by the available computational technology which relied for the most part on Hollerith cards for data input. The 80 column format of the Hollerith card forced decisiveness in coding decisions onto the researcher. Even when the Hollerith card went the way of the Dodo, the constraints they imposed on computer-aided historical research lived on in database and analytical software whose »flatfile« appearance was predicated on its punch-card predecessor. The process of coding occupational data for analysis thus remained for some time an exercise which could only be driven by rigorously defined questions. The aims of problem-oriented social history and the limitations of the available technologies seemed in this one respect to cross-fertilize.(6)

The marriage was not, altogether a happy one. Leaving aside for the moment the strait-jacket that the computer imposed on historical data, problem-oriented history threatened to overlook through its directed gaze the rich texture of the historical record and consequently of the past.(7) It also failed to some extent to provide the means for subsequent researchers to verify its findings or even to draw comparisons with other similar studies. Data once coded were lost forever to the secondary analyst whose assumptions and research interests might very well be different than those of the original author.

The advent and widespread use of hierarchical and relational databases from the mid-1970s finally allowed historians to break free of the flatfile format and to preserve data as more or less as they were found in the original source.(8) Managed in this fashion, any one dataset could support multiple coding schemes and consequently any set of interests or assumptions. New database technology cannot, however, overcome the fundamental problems involved in coding occupational data. The exercise is still by and large a subjective one and for large projects it is time consuming and expensive as well. The millennial claims made by the gurus of relational and hierarchical databases in advocating their techniques stress the technical capacity to support multiple interpretations of data; they rarely address the cost involved in actually implementing them. Coding tens of thousands of records for analytical purposes is quite simply an expensive

operation which most research projects will only want to contemplate doing once. Despite considerable technological advance, the problems of coding occupational information are still left largely unsolved. The questions as to how to derive workable occupational coding schemes which take account of the subjective and contestable nature of the exercise is as potent today as it was in the late nineteenth century. The remainder of this article not attempt a solution to the problems involved in coding occupational data. No one solution can possibly exist to satisfy historical investigations ranging across historical time and space. It will instead propose a framework or a scaffolding with which solutions to individual or research specific coding problems may be worked out.

The framework is developed from experience of a large-scale investigation into the social origins university, experiences and career destinations of 20th-century members of Oxford University.(9) Although the coding scheme adopted by the study to manage the wealth of occupational data that was gathered is perhaps specific to the analysis of a twentieth-century educated elite, it is hoped that the approach to the problem of occupational coding that has been adopted may prove generally useful.

Briefly, the study aimed to collect biographical information on a 10% sample of Oxford's members, 1900-67 or nearly 12,000 people. Data was collected through record search for Oxford members 1900-39, and through survey questionnaire for members 1940-67. The two different means of data collection were determined firstly by the demographic characteristics of the population. The 1900-39 cohort would not, for obvious reasons, provide a sufficiently large response rate to justify the expense of a postal survey (the project was eventually extended to survey university members who came to Oxford in the 1930s and in the period 1970-9). The adoption of different techniques of data collection was also influenced by a survey of the extant published record. This indicated that a wealth of information on students' social, geographical and educational backgrounds was available in university registers throughout the period 1900-67. Although these data were not as precise as was desirable (occasionally members used the term »businessman« in describing their fathers' occupation), they were at least comprehensive. Information on students' career destinations was not quite so accessible. Preliminary investigations indicated that the published record would supply careers information for approximately 60% of the students who attended Oxford through the 1930s.

The fact that the published record became less comprehensive after 1939 is itself indicative of the changing pool from which the university recruited its students, and of the changing pattern of students' eventual careers. Before WWII the lion's share of Oxford's men graduates went into self-documenting professions (eg medicine, the clergy, and the law) whose registers could be plundered systematically.(10) Furthermore, a very high

proportion of Oxford men in this period came from so-called public (private in any other context) schools, many of which kept up extensive biographical records as a means of fostering the allegiance (and ultimately benefactions) of their old members.⁽¹¹⁾ Many of the Oxford colleges to which all university members were affiliated, were for similar reasons as assiduous in keeping track of their old members. After 1939, record search offered increasingly diminishing returns. Students were, quite simply, entering into a wider variety of careers, branching out especially into industry and commerce where their presence was unlikely to be identified in the registers of professional associations. Moreover, the upkeep of public school and college registers fell into abeyance. Postal surveys therefore became essential for data collection about people in that cohort for which they were feasible.

Data collection through record search and survey produced biographical information on 11,980 individuals and yielded approximately 75,000 distinct occupational records pertaining to students and to their fathers, mothers and spouses.⁽¹²⁾ The sheer amount of occupational data collected by the project highlights the first important principle of preparing data for analysis. That is, coding must aim first and foremost at bringing the available data to bear on the intellectual problems that the researcher sets out initially to solve. Though the recommendation seems common place it is fundamental. Given the shoe-string budgets of much current research in history (or historical sociology), a coding scheme which ultimately proves ineffective can be a costly mistake which may make the difference between completion and non-completion of a project.

A corollary of this first fundamental principle is that coding schemes must be based on rigorous formulation of the problems involved in investigation and on extensive pilot projects. It may be safely argued that British twentieth-century university education in general and an Oxford education in particular was in this period the province of sons and daughters of the middle class, who are on their way into middle-class careers. The sons and daughters of the landed gentry are barely in evidence amongst Oxford's students after WWI. At the other extreme, men and women from working-class backgrounds never made up more than 12% of the student population at any one time and a far smaller proportion wind up in working class jobs after graduating. The coding scheme used to prepare for analysis the data on fathers' and students' jobs needs, therefore, to reflect the orientation of the project around inter-generational middleclass career mobility. Consequently, widely accepted occupational classifications used in analyses of a genuine crosssection of the population are of little use.⁽¹³⁾ With such a scheme the vast majority of Oxford students would fall unhelpfully in a categorical range so narrow as to defy any indication of change over time in either the university's social composition or its func-

tion in recruiting to various occupations. The ramifications of the specific problem-oriented approach are far reaching as they seem to suggest that attempts in other humanities disciplines to develop appropriate standards for treating machine readable data may have limited application to historical investigation.(14)

Having established the socio-economic range which a coding scheme must cover is not in and of itself sufficient. Meaningful stratification must also be settled with respect to the questions whose answers are sought in part from the data. Here, the project was influenced principally by a growing literature which attempts to establish twentieth-century universities' relationship with the learned professions (eg the clergy, medicine, the law, and teaching); its ability to meet industry's demands for competent management, and especially after WWII for scientists, technologists and other highly trained specialists. With particular reference to Oxford, the university's nineteenth century elitist heritage also highlighted the importance of examining whether inter-generational succession advanced or declined in the course of the twentieth century.(15)

Questions involving the function of a university in recruiting to various careers tended to underline the importance of a coding scheme which could support both sectoral analyses as well as analyses aimed at determining movement along a hierarchically arranged socio-economic scale. The latter was necessary to establish in factorial analyses whether university education had any impact on social mobility. It was also important to determine the »value added« by a university education. Only a coding scheme providing for some hierarchical analysis could enable the researchers to determine the proportion and composition of university-educated people who took jobs for which university education was, in fact, required.

Finally, the coding scheme needed to support analyses of Oxford role as a regional, national and international university. The twentieth century witnessed a dramatic increase in the number of British universities, in the amount of state funding for and centralized control of higher education, and the evolution of national school-leaving examinations. In short, the past 90 years have seen the development of something approximating a national »system« of higher education.(16) Oxford's place within that system was measurable in part with reference to the geographical origins of its recruits. Comparing students' geographical origins and the regional setting of their postuniversity careers promised to define the university's national and international role more precisely. Such analyses could determine whether Oxford returned its recruits to their region of origin or simply channelled them into London and the southeast thereby draining other areas of educated talent? They also promised to demonstrate whether the university's international clientele changed significantly through the years of imperial decline.

The problem-orientation of the research, then, framed the broad outlines of the occupational coding scheme that was adopted. To accommodate three related analytical aims - sectoral, hierarchical, and geographical - a tripartite coding scheme was produced. According to the scheme (set out in part in appendix) each occupational record was coded in three categories which could then be treated in analysis either independently or in conjunction. The first category identifies economic sectors and includes (among others) the armed forces industry, commerce, finance, public service, the clergy and church related activities, charitable and social work, education, land ownership and farming, health care, and the legal profession.

A second category supports analyses along hierarchical lines. Such analyses are, however, strictly limited in their application by the nature of the data that were available to the project. Clearly the information most desirable for analyses of socio-economic stratification bears on individuals' income and wealth. Such data simply could not be found through record search for such a large number of people without a vast amount of time and money invested in plundering tax rolls and probate inventories. Information on income was obtainable through survey questionnaire, but its benefits had to be weighed up against the costs of a diminished response rates caused by survey questions which were deemed too personal by survey recipients. In general, the more personal and prying the questions are on a survey, the fewer people will be bothered to respond to it. Moreover, there was some evidence that including questions concerning on would discourage recipients from low income brackets from responding to the survey. There was some danger then of skewing results toward the most successful Oxford members.

The process of weighing up the costs and benefits of collecting information on wealth for only about half of our sample (the surveyed population) against the risk of upsetting the representativeness of our results is evident in the occupational coding scheme. There it will be seen that the hierarchies are based not on income (a question excluded from the survey questionnaire) but on the nature of the work involved in a particular job. The secondary occupational category thus comprises a five-tiered hierarchical guide and includes categories for top executive and proprietary personnel, for administrators and managers, for skilled workers, routine non-manual workers, and for manual workers.(17)

There are also categories in the second level of the coding scheme which are specific to first-level or sectoral codes indicating areas of specialism. For example, teachers in the educational sector are stratified according to the kind of school in which they were teaching. The scheme provides for teachers in adult education, and at teacher training colleges, universities, and secondary and primary schools. It also indicates positions within the

hierarchy which is specific to the profession. Headmasters and and university professors, for example, are distinguished from secondary teachers and university lecturers respectively. Finally a third-level code indicates for each occupational record the geographical location of the job that is represented.

Pilot investigation was as influential as problem orientation in shaping the coding scheme as it established more precisely the kinds of data the project was likely to encounter. Pilot studies involving data collection for a small number of people (100 each by record search and survey) showed that any analysis have to take account of the fact that data gathered through record search are less specific than those gathered through postal survey. Time series analyses which draw on the data gathered through both methods require more general treatment to take account of the ambiguous occupational descriptions such as »businessman«, »merchant« and managing director« which crop up in the record-search data. At the same time it is desirable that analyses of the post-1939 data gathered through survey questionnaire be able to exploit the precise occupational descriptions that the questionnaire was designed to invoke. The problem is less acute in the secondary and tertiary (hierarchical and geographical) categories. These can either be completed or indicated as unknown where appropriate.

More problematic is the sectoral category. Here, it proved necessary to use broad sectoral codes which can be sub-divided where appropriate on the basis of the more precise survey data. Thus, the category for the armed services accommodates the vague job description »military service« occasionally encountered through record search. Meanwhile, various sub-divisions of the category accommodate the more specific information about branch and type of military service invoked through survey, eg »captain, Indian army«. Data in more refined sub-divisions of any particular category can then be considered as belonging in the more general category where time-series analyses require.

Lessons drawn from pilot investigation, then, point to another feature of a framework for occupational coding; that is the benefits of a scheme which rests on divisible categories. Such a scheme is also important because it provides a degree of flexibility in analysis. Analyses of prosopographical data are necessarily regressive. Questions initially asked of the assembled data produce results usually in the form of statistical tabulations which help the researcher to further refine the questions or indicate new avenues of investigation hitherto not thought of. Even the most thoroughly documented and rigorously conceptualized research strategy needs to provide for this kind of interactive interrogation. Evidence that the problem is a general one inherent in much sociological and quantitative historical research is available in the flexible re-coding mechanisms that are available in most standard statistical software. But no re-coding facility

will appear terribly »user friendly« if after initial interrogation the researcher is forced to reconsider the data at its most atomic level eg, by recoding the raw data. Instead, the researcher wants to be able to flexibly recombine the data as they are already coded to produce meaningful analytical categories.

A few figures from the Oxford project are sufficient to make the point. Coding 75,000 occupational records required the equivalent of one person working full-time for over six months. In a project of this scale, it is therefore imperative to ensure that the coding scheme supports recoding at a level which is at least one step removed from the raw data. With the current scheme the same 75,000 discrete job records yielded only 700 distinctive combinations of sectoral and hierarchical codes (when geographical codes are considered, of course, the number of distinctive code combinations increases considerably). Re-combining the 700 distinctive code combinations to create new analytical categories is a job which takes at most a long morning; and it is at this level in the data that any re-coding should take place. With the divisible categories outlined above, it is possible to construct an analytical category of »educationalists«, for example, by looking to the 700 distinctive code combinations for jobs involving education-related work from across the various sectors (eg where an educational-related activity is indicated for jobs in the educational, local government, sector, church-related and legal sectors, and in local government).

Providing divisible categories in occupational coding schemes also ensures a degree of flexibility in expanding the scheme to make it more comprehensive or to take account of new or changed assumptions about the nature and meaning of the occupational data themselves. Should it prove desirable, for example, to create a sectoral category for jobs in extractive industries (eg mining, oil companies), only the discrete records already coded as being in the »industry« category need to be addressed at the atomic level. Nowhere should the creation of a new sectoral category involve looking through raw data currently resident in several different sectoral categories. The same would prove true for refinement of the second-level or hierarchical codes. The scheme will, for example, be extended to support hierarchical analysis of civil servants when research into the changing structure of the civil service is completed 18) Only the discrete job records already coded in the civil service sectors need to be examined and re-coded.

The flexibility of the coding scheme is also important if the data are to have much value for secondary and comparative analysis. In theory, a host of different research assumptions should be sustainable by constructing new analytical schemes from the approximately 700 unique combinations of coded categories. Never should the secondary analyst have to re-code the entire dataset of 75,000 occupational records. The scheme holds forth si-

milar advantages for comparative study as the 700 unique code combinations may be re-combined to replicate the analytical schemes used in other similar or comparable studies. In this respect, schemes developed along the lines suggested above may fulfill the aims of standardization without enforcing rigidly defined standards on researchers involved in fundamentally different investigations.

Finally, the key to the scheme's flexibility is the atomic nature of the categories used in its construction. That is, all of the codes in each of the three categories are rigorously defined so that they represent a distinctive data type. Only those data which meet the criteria devised for the particular data type are entered into that category.⁽¹⁹⁾ The sectoral code of health care, for example, comprises only those occupational records which indicate a job in a hospital, nursing home or other related institutions, or in a private or public general medical practice. By this criterion, medics in the armed forces are excluded from the health care sector (though they may be treated alongside hospital doctors and general practitioners when analytical categories are devised from the 700 unique code combinations). So are school nurses who would appear in the educational sector with the second-level code for nurse.

The use of atomic categories is important for two reasons. Firstly, it allows the assumptions built into the coding scheme to be imparted to others who may want to use it for analysis or for constructive criticism of research results. It is also important if the scheme is to be implemented consistently across all records. In large projects where several people are engaged in coding occupational records, rigorously defined coding criteria are essential if consistency is to be achieved. In the ideal scheme, code categories would be so well defined that two people independently coding the same set of occupational records would agree 100% of the time. Even where projects are small enough so as to allow one person to code all the data in a relatively brief period of time, rigorously defined criteria ensure that the same assumptions are employed in coding from record to record, from hour to hour, and from day to day. One way of testing the consistency of any coding scheme is to see to what extent different people who are apprised of the same set of rules can agree on how to code a given body of data.

The history and sociology of twentieth-century higher education is currently a growth area and it is hoped that the scheme outlined in appendix will be of some use to those actively engaged in it. For the most part, however, its usefulness lies in the framework for occupational coding that it reflects. No coding scheme can hope to overcome the problems thrown up by ambiguous or incomplete data, or by the essentially contestable assumptions from which they derive. A systematically contrived scheme can, however, document for the secondary analyst and critic where assumptions

were introduced into the analysis so that they may be undone or simply treated differently. Further, it is proposed that such a scheme will serve the purposes of both the critic and the secondary analyst without either having to return to the raw data for recoding. The development of such a scheme relies less on the application of standardized codes as on the application of standard procedures for coding. The paper, in other words, offers a meta-standard or framework whereby researchers may at least agree the rules of the game they are playing if not the results of each and every match.

Briefly, that framework suggests that occupational coding schemes are by necessity developed with reference to the intellectual orientation of the research at hand and to pilot investigations which demonstrate the strengths and weaknesses of the data that are likely to be collected. Where such an approach tends to tailor the coding scheme closely to narrowly defined research aims, the introduction of strictly defined but divisible coding categories will support more general usage. Divisible categories offer additional advantage to the researcher by protecting against having to return to the raw data where the re-combination of coded occupations is required in analysis. When the raw data do have to be invoked, for example, to sub-divide categories already established, they may be called up in limited sets providing that initially established categories are rigorously defined according to documentable and consistent criteria. The use of rigorously defined individual codes also ensures consistency when job records are being coded by several researchers or by any one researcher over any period of time. This framework cannot, alas, remove the frustration involved in the thousands of decisions required to satisfactorily place unique job records into their appropriate category. It may, however, ensure that the effort pays a greater dividend for individual research projects, their critics, and their future users.

Appendix

The occupational coding scheme is set out below under two heads reflecting sectoral and hierarchical codes assigned to each occupational record. The geographical component of the scheme is omitted except where it appears in examples of individual codes (C for commonwealth and O for other overseas). The actual geographical scheme uses pre-1973 British counties to describe British addresses, and the names of the relevant foreign country for non-British addresses.

In the first section on sectoral codes, the over-arching sectors (eg industry) are enumerated and presented in capital letters. Sectors within the divisible overarching category are set out beneath in no order of importance. Beneath each sectoral code, examples are given showing how

sectoral, hierarchical and geographical codes may be used in combination (commas separate the three codes). Those sectoral categories indicated with an asterisk will normally be susceptible to hierarchical categorization which indicates proprietary and executive status, other management, routine white-collar, skilled manual and semiand un-skilled manual work. Those sectoral codes for which comprehensive hierarchical categories have yet to be worked out are also indicated.

I. Sectoral Code

1. MILITARY/POLICE ESTABLISHMENT

A armed forces (eg army, airforce, navy - largely unstratified)

A,C	army chaplain
A,M	army medic
A,RES	naval scientific officer
A,VET	army vetirenary surgeon
A,,C	Indian army
A,,0	US Air Force

AC* civilian auxiliary to the armed forces (naval dockyards etc)

AC,D	naval stores officer
AC,W	fitter in naval dockyard

AP fire service, police and prison services

AP,Z	police commissioner
------	---------------------

2. COMMERCE

ADVT* advertising firms

ADVT,DES	designer, draftsman
ADVTJ	copywriter

CONS* management consultancy

LEIS* arts, entertainment and sport

LEIS,A	actor, artist, author, conductor, musician
LEIS,B	bookmaker, courier, film company agent
LEIS,D	film company manager
LEISJ	film company director, editor, producer

M* (press, radio, television)

M,J	journalist, BBC reporter
M,Z	newspaper proprietor

MKT*	market research	
MKT,WN	interviewer	
O*	other commerce not taken account of in sub-divisions in this section	
O,B	import broker, wholesale merchant	
OC*	colonial/commonwealth trade	
OC,B	tea broker	
OC,Z	coffee merchant (proprietor or top executive)	
OR*	retail commerce (high street outlets and retail trade)	
OR,B	buyer for Marks & Spencer	
OR,N	sales assistance at Woolworth's	
OR,SM	baker, butcher, small shopkeeper, newsagent	
OR A*	retail of art or precious works, gems etc	
TRANS*	transportation	
TRANS,D	controller, British-Rail stationmaster	
TRANS,N	dockyard worker, longshoreman	
TRANS,PI	airline pilot	
TRANS,Z	shipowner	

3. INDUSTRY

B*	book publishing	
B,B	literary agent	
B,J	editor	
B,WN	proofreader	
BP*	printing	
BP,N	machine operator, printer's reader	
BP,W	skilled print worker	
BP,WN	clerical assistant	
BU*	building	
BU,CONTR	building contractor	
BU,DES	draughtsman	
BU,ENGR	civil engineer	
BU,N	building worker	
BU,R	architect/surveyor for building firm	
BU,W	master electrician	
COMP*	computer industry	
COMP,PRO	computer programmer	
I*	other industry not accounted for in other sub-divisions	
I,W	chargehand, fitter, foreman, lathe operator	

I,W	industrial labourer	
I,Z	secretary/ chief executive/ owner	industrial company
IF* food industry		
IF,B	tea taster	
IF,D	dairy manager	
IF,N	brewery worker	
TU* trade union		
TU,D	trade union official	
UTIL* utility companies		
UTIL,D	manager, gas board	
UTIL,N	water works, worker	
UTIL,RES	scientist, UK Atomic Energy Authority	

4. PUBLIC SERVICE

CCS colonial civil service (largely unstratified)

CES colonial educational service (largely unstratified)

CMS colonial medical service (largely unstratified)

CS home civil service (largely unstratified)

CSJNSP tax inspector

FO foreign office (largely unstratified)

GPO* general post office including British Telecom

G PO ,D head postmaster

GPO,N postman

GPO,W telephone linesman

ICS Indian civil service (largely unstratified)

ICSJVIAG magistrate, ICS

IES Indian educational service (largely unstratified)

IMS Indian medical service (largely unstratified)

LG local government (largely unstratified)

LG,M local government medical officer

LG,R municipal surveyor

LG,WN planning assistant

LGE local government education authorities

LG,D chief education officer

LG,Z director of education

LGSW local government social services

LGSW,D senior probation officer

PB* public body eg Citizens' Advice Bureau, English Tourist Board, National Trust

V elected political position given as occupation (largely unstratified)

5. CHURCH RELATED AND RELIGIOUS

CH* church related and religious

CH,C cleric, deaconess, nun

C H,WN lay preacher

CH,Z church commissioner

6. MISCELLANEOUS

CRT* charitable organizations

CRT,D Oxfam manager

CRT,VOL voluntary charity worker

CRT,Z YMCA national secretary

IO* international organization

IO,D United nations official

NC not codeable

NK not known

NL unemployed

OI* used for unspecified business/industry (eg where information provides simply »businessman« or »company director«)

SO* purely social organizations eg Women's Institutes, English Speaking Union

VOL* voluntary work

VOL,LIBR voluntary librarian

VOL,SW voluntary social worker

7. EDUCATIONAL (see section II for second-level codes largely specific to educational sector)

E* educational

E,D university registrar

E,S school teacher

EJRES social science research officer

EJLIBR archivist/ librarian

ES* research institute

ES,RES

Medical Research Council

ES,Z

director of British Council

8. FARMING AND FARMING-RELATED, ESTATE MANAGEMENT,
LAND OWNERSHIP

GF* farming and landowning

GF,B

land agent

GF,W

cowhand, gardener

GF,Z

GFP* planter involved with tea, coffee, rubber

GP* non-farm property

GP,AU

land auctioneer

GP,B

estate agent

9. HEALTH CARE

HC* Health care (see section II for second-level codes largely specific to health care)

HC,D

hospital administrator

HC,RES

medical microbiologist

HC,M

hospital consultant physician

10. SERVICE INDUSTRY

HO* hotel, restaurant and catering

HO,N

butler at a college, chauffer, waiter

HO,SM

guesthouse keeper

HO,Z

proprietor of a large hotel or hotel chain

HO,W

caterer/chef

11. FINANCE

K* finance

K,B

insurance agent/broker, stockbroker

K,D

bank official

K,T

chartered accountant

K,WN

bank clerk

12. LAW* legal profession (see section II for second-level codes largely specific to law sector)

LAW,D

magistrate's officer, registrar in probate court

LAWJL	lawyer
LAW,MAG	magistrate
LAW,WN	solicitor's secretary

II. Hierarchical code indicating specialism where appropriate

A. Hierarchies

Z	executive/proprietary, chairperson, chartered secretary, company secretary, general secretary etc
D	administrative/managerial, administrator, assistant manager, chief of department, etc
B	sales manager (to be considered on a par with D above)
WN	routine non-manual: administrative assistant, cashier, clerk, supervisor
W	skilled- and semi-skilled manual
N	unskilled manual B. Codes indicating specialism
For specific use with educational sectoral code	
AO	adult education
EX	examiner
O	other kind of teacher not otherwise specified in other codes (eg language schools)
P	coach, private tutor
S	secondary and primary school teacher
SZ	secondary and primary school headmaster/mistress
TT	teacher training college
U	university/college teacher other than professor
UP	university/college professor
UZ	university/college head, chancellor, president, principal etc
For specific use with church and religion related sector	
C	ordained minister
CS	cleric and primary/secondary school teacher
CSZ	cleric and primary/secondary school headmaster/mistress
CU	cleric and university/college teacher
CUP	cleric and university/college professor
CUZ	cleric and university/college head, principal etc
CZ	bishop, cardinal etc
MISS	missionary
Used	largely with health care sector HC
F	other health related profession not otherwise specified eg osteopath, psychologist

M	physician
MC	consultant physician
MU	medic and university/college lecturer
MUP	medic and university/college professor
NU	nurse
OPT	optician
Used	largely with the legal sector LAW
L	lawyer (including barrister and solicitor)
LJ	judge
LU	lawyer and university/college lecturer
LUP	lawyer and university/college professor
MAG	magistrate
Other codes	indicating specialism
A	author, entertainer, musician, writer
ADVT	advertising
AU	auctioneer
CHEM	chemist
CONS	adviser, consultant
CONT	contractor
DES	designer, draughtsman
ENGR	engineer
FIN	fine craftsman eg optical instrument maker, watchmaker
FS	forestry
INSP	inspector, examiner
J	editor, journalist, media producer
LIBR	archivist, librarian
PHAR	pharmacist
PRO	computer programmer/analyst
PS	personal secretary
R	architect, surveyor
REA	research/specialist in the arts/humanities eg art historian
REB	research/specialist in the social sciences eg economist
RES	research/specialists in the sciences eg microbiologist
ST	further study, trainee
SW	social/ welfare worker
T	accountant
TR	treasurer
TRL	abstractor, translator
VOL	voluntary work

Notes

* The work on which the article is based was made possible through the generous support of the University of Oxford, and by grants from the Levehulme Foundation, and the Economic and Social Research Council. It could neither have been conducted nor refined without the herculean efforts of Dr Mark Pottle and Ms Sinead Smith who spent quite literally months implementing and expanding the occupational coding scheme herein described. Whatever shortcomings that remain with the scheme are, however, the sole responsibility of the author.

- (1) D.V. Glass and J.R. Hall, *Social Mobility in Britain* (London, 1954).
- (2) For a review of Lazarfeld's impact on American sociology see Allen H. Barton, »Paul Lazarfeld and Applied Social Research«, *Social Science History* 3:3(1979)
- (3) Lawrence Stone, »History and the Social Sciences in the Twentieth Century«, L.Stone, ed., *The Past and the Present* (London, 1981); P. Abrams, »History, Sociology, Historical Sociology«, *Past and Present* 87(1980)
- (4) J.J. McCusker's review of T.K. Rabb, *Enterprise and Empire: Merchant and Gentry Investments in the Expansion of England, 1575-1630* (Cambridge, Mass., 1967) in *Historical Methods Newsletter* 2:3(1969); Michael B. Katz, »occupational Classification in History«, *Journal of Interdisciplinary History* 3:1(1972), W.A. Armstrong, »Appendix D. The Classification of Occupations«, E.A. Wrigley et al., eds., *An Introduction to English Historical Demography* (London, 1966), and »The Use of Information About Occupations«, in E.A. Wrigley, ed., *Nineteenth Century Society: Essays in the Use of Quantitative Methods for the Study of Social Data* (Cambridge, 1972)
- (5) See for example Francois Furet, »From Narrative History to Problem Oriented History«, Francois Furet, ed., *In the Workshop of History* (1984), chapter 3. The article first appeared in 1975.
- (6) Louis T. Milic, »The Next Step«, *Computers and the Humanities* 1:1(1966), 3-6.
- (7) Arguments for and against this position have been rehearsed ad infinitum. See, for example, William O. Aydelotte, »Quantification in History«, *American Historical Review* 71(1966) 803-25; or Robert Fogel and G.R. Elton, *Which Road to the Past: Two Views of History* (London, 1983)
- (8) Daniel I. Greenstein, »A Source-Oriented Approach to History and Computing: The Relational Database«, *Historical Social Research* 14:3(1989).

- (9) Work on the project has been made possible with the assistance of grants from the Leverhulme Foundation and from the Economic and Social Research Centre.
- (10) Preliminary results from the project are available in Daniel I. Greenstein, »Oxford and Social Class in the Twentieth Century«, B.H. Harrison, ed., *The History of the University of Oxford: The Twentieth Century*, vol. 8. (Oxford, forthcoming) or through direct correspondence with the author.
- (11) Richard Holt, *Sport and the British: A Modern History* (Oxford, 1989), 114 offers a slightly different view. School registers, associations and the old school or college tie allowed British elites to take their status with them on their travels to the distant posts of empire.
- (12) Record search produced information on students' careers and on jobs held by their fathers when students first came to Oxford. The survey questionnaire provided vastly more, and more structured information on the jobs students held immediately after Oxford and every 10 years thereafter, on jobs held by fathers' and mothers' at the time the student came to Oxford, and on the jobs held by students' spouses at marriage and at subsequent intervals.
- (13) Obvious candidates include the schemes used by the Oxford Social Mobility Study, in the compilation of the British census, by the Population Investigation Committee or by Hall's and Caradog-Jones's Social Survey. For a fuller discussion of these see M. Hall and Caradog Jones, »The Social Grading of Occupations«, *British Journal of Sociology* 1 (March 1950); John H. Goldthorpe and Keith Hope, *The Social Gradings of Occupations: A New Approach and Scale* (Oxford, 1974); John H. Goldthorpe et al, *Social Mobility & Class Structure in Modern Britain* (London, 1980), 39²
- (14) The Text Encoding Initiative currently underway is one such example. See ACH-ACL-ALLC, *Guidelines for the Encoding and Interchange of Machine Readable Texts*, ed. by CM. Sperberg-McQueen & Lou Burnard, Draft Version 1.0 (Chicago and Oxford, 1990)
- (15) See for example M.J. Weiner, *English Culture and the Decline of the Industrial Spirit, 1850-1980* (Cambridge, 1981); Michael Sanderson, *The Universities and British Industry, 1850-1970* (London, 1972).
- (16) Roy Lowe, *Education in the Post-War Years: A Social History* (London, 1988); For a European-wide perspective see K. Jarausch, ed., *The Transformation of Higher Learning, 1860-1930* (Chicago, 1983)
- (17) This aspect of the scheme is broadly based on that used by Hall and Caradog Jones op cit
- (18) The same applies to the sectoral code for local government which will be stratified after research into the local government service.
- (19) The concept of the atomic coding category shares much in common with that used to define fields or data types in most standard DBMS.